CAS LX 500 Lg Acq and Ling Thy
Fall 2013, September 24

7. Experimental design, the DE in L2

1 The point of doing an experiment

Why do experiments?

The point of doing an experiment

One does an experiment to test a hypothesis.

This might seem obvious. But a good experiment is one that helps determine whether a hypothesis one has about the Way Things Work is correct or not. That is, you *start* with an understanding of a sort, and you look at what kinds of predictions this makes. What would the world look like if this were true? The experiment is done to help you decide whether the world in fact looks that way.

Belaboring the point

"Fishing" is not generally good in the context of an experiment. An experiment should be carefully designed so that the results separate two possible hypotheses.

For example, I you wonder to yourself "Do children acquiring English use the verb *kick* before they use the verb *hug*?" and then set out to find out by doing a corpus search on CHILDES, you'll find out the answer. But you won't know anything else. Is it surprising that they do or don't? What does it tell us about the process of language acquisition? The answer might be something like "Yes, (certain?) children acquiring English use the verb *kick* earlier than *hug*." Yes, and. . . ?

A slight defense of fishing

However, that being said, if you find that—after looking at a lot of children's transcripts—children consistently use the verb *kick* first, you do have a data point. It's not really informative by itself, but it can be used to *generate* a hypothesis. The question you would then ask is: *Why?* What could the reason be? Does it have to do with the input? Does it have to do with the phonology? Does it have to do with the semantics?

Fishing is not where investigation stops, but it is sometimes useful in getting started. (Note, though: experimental purists might dispute even this—wanting the entire hypothesis spelled out before looking at the data.)

Hypothesis

Suppose it has to do with the phonology. What about the phonology? Well, maybe it has to do with the difference between *h* and *k*—maybe *k* is "easier" to produce earlier.

Is that true? Well *if it were true*, we'd expect to find a lot of words beginning with *k* early on, and not a lot of words beginning with *h*. So, then, the real experiment can begin. We look at the early transcripts, count up the words with *h* and *k*, and see. Suppose there *are* a lot of *k* words relative to *h* words. Has it been shown that *k* words are acquired earlier? What makes *k* better than *h*? Are they *trying* to say a lot of *k* words? Etc. etc.

2 The Definiteness Effect

2.1 The DE in English

Acknowledgments re: the DE experiment

This is ongoing work done in collaboration with Alyona Belikova, Öner Özçelik, and Lydia White (McGill University) and Tanja Kupisch (University of Hamburg). The main results reported here appear in White et al. (2009).

This research was funded by grants from the Social Sciences and Humanities Research Council of Canada (SSHRCC) and from the Fonds québécois de la recherche sur la societé et la culture (FQRSC).

There-constructions

Many sentences like (1a) can be rephrased using *there*, as in (1b).

(1) a. Several flies were in my soup.
b. There were several flies in my soup.

There are some restrictions, however.

***There*-constructions: Predicate type**

The predicate must be "stage-level"—that is, it must represent a temporary property.

(2) a. Several flies were eating my soup.
b. There were several flies eating my soup.

(3) a. Several flies were unusually brave.
b. * There were several flies unusually brave.

***There*-constructions: Subject type**

One type of subject works (e.g., *several X*). Another type doesn't (e.g., *every X*).

(4) a. Several flies were eating my soup.
b. Every fly was eating my soup.

(5) a. There were several flies eating my soup.
b. * There was every fly eating my soup.

The Definiteness Effect ("DE")

Definiteness Restriction

In a *there*-construction, the associate cannot be strong.

Definite NPs are in in the second category (called "strong"). Indefinite NPs are in the first category (called "weak"). Here is a bigger list (Milsark 1977).

Weak

a, *some*, *many*, *few*, cardinal numbers (*one*, *two*), bare plurals (*seats*, *votes*), mass nouns (*cheese*)

Strong

the, *all*, *most*, *every*, demonstratives (*this*, *that*), possessives (*my*, *his*), personal pronouns (*I*, *you*), proper names (*Norm*, *Al*)

At least for the moment, it doesn't matter what exactly the property is that differentiates "strong" and "weak" in this context. Enç (1991) argues that it is actually a distinction of specificity, not of definiteness. Perhaps this might even differ between languages akin to the way Ionin (2006) proposes article systems vary, although this should not be the starting assumption. The main thing is: there is some property (which we assume is the same for Turkish and for English).

Definiteness effect: Exceptions

The subtlety: It is not as simple as ruling out "there is the…" however—this word sequence is acceptable in the right context.

(6) *Deixis*:
Oh, look. **There's the** missing soup ladle.

(7) *List readings*:
What should I read first?
There's the article on determiners, and there's the *Harry Potter* book...

2.2 The DE in Turkish

Turkish also shows the DE

We see definiteness effects in Turkish as well (Enç 1991)—it's not English-specific.

(8) a. Bahçe-de birkaç çocuk var.
garden-LOC some child exist
'There are some children in the garden.'

b. * Bahçe-de o çocuk var.
garden-LOC that child exist
('There is that child in the garden.')

Negative existentials in Turkish

But it is not precisely the same. Notably, in the *negative*, the DE disappears.

(9) a. * Bahçe-de o çocuk var.
garden-LOC that child exist
('There is that child in the garden.')

b. Bahçe-de o çocuk yok.
garden-LOC that child not-exist
'There isn't that child in the garden.'

Articles in Turkish

Another difference between English and Turkish is that Turkish does not have a definite article (Kornfilt 1997).

(10) a. Ali kitab-ı okudu
Ali book-ACC read
'Ali read the book.'

b. Ali bir kitab-ı okudu
Ali a/one book-ACC read
'Ali read a book.'

c. Adam gel-di
man come-PAST
'The/a man came.'

3 Experimental investigation of the DE in L2A

3.1 Motivating the experiment

Poverty of the stimulus wrt DE?

Transfer The DE operates slightly differently in English and in Turkish. Specifically, in negative existentials definites are allowed in Turkish and disallowed in English. This allows us a way to detect L1 transfer.

Instruction The pattern is subtle—it involves not just *the* and *a*, but the various other weak and strong quantifiers. The pattern also has exceptions (list, deictic), so a simple rule—even if it were taught—would lead to non-English-like judgments.

Frequency It also can't be just that *There is the* is rare—it's only allowed in the exceptional contexts, and it is ultimately *more* frequent than the grammatical *There is one*.

Previous results on DE in L2 English (production)

- White (2003): Production data from one advanced endstate Turkish speaker.
- Lardiere (2005): Production data from one advanced endstate Mandarin speaker.
- White (2008): Elicited production (picture description) from 18 Turkish speakers and 15 Mandarin speakers, various proficiency levels

The results...
No DE violations were found.

Methodological questions about production data

- Production data can't distinguish between *ungrammatical* and *dispreferred*. But we want to know about grammaticality. (Some subjects didn't produce *there*-insertion at all. Should that really count as obeying the definiteness restriction?)
- In a production task, one has little control over the context.

- Certain sentence types did not occur in production data (i.e. negative existentials—the crucial test case for transfer).

To get at the questions more directly…
We designed a *judgment* task.

3.2 Defining the questions

Defining the big question

So, now, there's an experiment to do. What are we trying to discover?

The DE is universal enough that it can be assumed to be part of UG (a principle, or derived from principles), although there is a specific way in which it differs between languages (a parameter, or derived from parameters).

Do L2'ers have access to the different parametric options?

Defining the small questions

How can we answer this question?

Do Turkish L1 speakers *transfer* the property that suspends the definiteness effect in negative existentials? That is, do they transfer the parameter setting of their L1?

Or, do they *acquire* the English version, where the definiteness effect holds in both positive and negative existentials?

(Or, do they simply act somewhat "randomly," differently for different combinations of words?)

Determining how to answer the questions

To determine whether this property is transferred, we need to look at the specific case where there is a difference between the parameter settings. That is, we need to know **whether they accept strong DPs in negative existentials**.

So, we test them on those: *There isn't John at home.*

Except that's not good enough. Suppose they accept strong DPs in negative existentials. Is it because they transferred the Turkish instantiation of the DE? Or is it because they simply don't observe the DE at all?

Is there a DE in the L2'ers English at all?

To ensure that the DE is observed in the interlanguage of these L2'ers, we also need to check the basic cases as well. So, we need:

- There is a... (good in English, Turkish)
- There is the... (bad in English, Turkish)
- There isn't a... (good in English, Turkish)
- There isn't the... (bad in English, good in Turkish)

Except that's not good enough. They probably were explicitly trained on *there is(n't) a* and *there is(n't) the*.

Is the DE about *the* and *a* or about DP strength?

The real DE is about DP strength. So, we need to try other strong/weak DPs apart from *a* and *the*, things that they aren't likely to have been trained on.

- There are many...
- There are most...
- There aren't many...
- There aren't most...

Except that's not good enough. Because, in certain circumstances, *there is the* is ok in English too.

If DE violations are accepted, are they exceptions?

The main exceptions are the list reading and the deictic reading.

Let's see I need to pack. What do I need to bring? Well, there's the book about cars from Greece. I need to remember to bring that.

Oh, and, look over there. There's the little box that lights up with different colors. I was wondering where I'd put that.

The only way to differentiate these exceptions from normal existential constructions is by setting up a context. So, we're going to need to provide contexts for these sentences.

Methodology

So, we'll provide a context for the sentence, and then have them judge the sentence. We'll have them say it is either "natural" or "unnatural."

Except that's not good enough. They might say that it's "unnatural," but because they don't know/like one of the words we used, for example.

If they say "There aren't most people in the elevator" is unnatural because it would be better, in their view, to say "There aren't most people in the lift," then they haven't actually rejected this on the basis of the DE, but for an irrelevant reason. We don't want to count that as a "reject." How can we ensure that? Well, we have them *fix* it.

3.3 Subjects and materials

Subjects and task

Subjects: Turkish-speaking learners of English. Proficiency: 3 low, 12 intermediate, 10 advanced. Tested in Istanbul.

Task: Contextualized acceptability judgments (n=90) on a computer screen, testing DE in affirmative and negative existentials, as well as deictic and list readings. (5 items per sub-category.)

Dependent vs. independent

What we are measuring in the experiment is the **acceptability judgment**. The model we have of what's happening is that the acceptability judgment is caused by several things: whether the subject is strong or weak, whether the sentence is affirmative or negative, whether the context supports a list reading or deictic reading.

So, the acceptability judgment is a **dependent variable**, depending on the various **independent variables** representing strong/weak, affirmative/negtive, etc. In the experiment, we vary the independent variables to see what the effect is on the dependent variable.

The dependent variable here is basically either "yes" or "no." (Not something like 1.73). This is called a **nominal** measure. We could have used an **interval** measure (rate this from 1–5).

Between- vs. within-subjects

We will also ask the same subject to respond to each of the different conditions. This is called a **within-subjects** design.

The primary alternative is a **between-subjects** design, where different subjects get different conditions. A between-subjects design is appropriate if, e.g., you are testing for the effects of a placebo vs. a drug over time. A subject either gets the drug or doesn't, you can't test the same subject's reaction in both circumstances.

Variables and levels

In statistical terminology, the dimensions along which conditions vary are called **variables** and the individual values are called **levels** (again, think drug and dosage).

Here, we have basically three variables: **Polarity** (levels: affirmative and negative), **Strength** (levels: strong and weak), and **Exceptional** (levels: exceptional and not). (Also, I guess: **proficiency**, levels: beginner, intermediate, advanced, native)

In the design, we want to balance things so that subjects see equal numbers of each level (in our case, 5 of each for the first two—the exceptional variable was only partially represented). The more the better, as far as statistics is concerned. In our experiment we also subdivided the "strength" levels into strong and definite, and weak and indefinite, and the "exceptional" level into deictic and list.

Controls

We also included certain **control** cases. These are generally there in order to rule out alternative interpretations of what the subjects are doing. For us, we had one set of sentences with indefinite subjects—to rule out the possibility that a subject simply doesn't allow indefinite subjects under any circumstances, and independently ungrammatical indefinites—to rule out the possibility that a subject simply rejects indefinites. These are basically analyzed separately, with the idea that if a subject does not respond correctly to the control sentences, the rest of the subject's data is not relevant to the question we're trying to ask.

Fillers

Another thing tests normally include are **fillers** (or "distractors"). These are sentences that are essentially ignored in the analysis, and are there primarily to

"fool" the subject. The point here is to avoid giving the test subject an idea of what you are testing for, to avoid having the subject start answering the questions not based on their intuitions but rather on some kind of strategy. For this purpose, the more fillers we have, the better. Minimally, probably half the items should be fillers—better, two-thirds. They should be comparable, but different.

In this experiment, we didn't really do this, although the control sentences served this purpose to some extent.

Balance

In general, the better balanced everything is the better. One thing to consider is how many of the items are expected to be judged "unacceptable." You don't want to have 80% of the items of the "unacceptable" sort. Best for this to try to keep the expected result about 50-50%. The fillers can help with this, if your design is such that most of the experimental items wind up on one side or the other, although it is best for the experimental items themselves to be balanced.

Other considerations

Along the same lines, you want to be sure that the items you are comparing are as close to the same as possible—that the main thing that differentiates them is the specific variable you are testing. So, you want to avoid superfluous differences, e.g., in word frequency or in plausibility, since these can also affect acceptability judgments. Also, for second-language speakers, particularly beginners, they may not have a big vocabulary.

That is: it's not good to compare "There isn't a capybara straddling the balustrade" and "There isn't the dog here."

Sentence types: Affirmative existentials

(11) a. Grammatical *there*-insertion with indefinites
There's a copy machine downstairs.

b. Ungrammatical *there*-insertion with definites
There's the only copy machine downstairs.

c. Grammatical *there*-insertion with weak determiners
There are several guests in the garden.

d. Ungrammatical *there*-insertion with strong determiners
There are your keys on your desk.

Sentence types: Negative existentials

(12) a. Grammatical *there*-insertion with indefinites
There doesn't seem to be a suitable bowl here.
b. Ungrammatical *there*-insertion with definites
There isn't the bowl here.
c. Grammatical *there*-insertion with weak determiners
There aren't many guests inside.
d. Ungrammatical *there*-insertion with strong determiners
There aren't most guests inside.

Sentence types: Controls

(13) a. Grammatical list contexts
Well, there's the bottle of wine from last night.
b. Grammatical deictic contexts
Look. There's the mouse again.
c. Grammatical indefinite subjects
A young woman is still waiting to see you.
d. Ungrammatical indefinites
A kitchen stove is broken.

Tricks

The creation of these items was actually relatively agonizing—the main problem was that we needed to be sure that the context made it clear that a list or deictic reading of the ones that are supposed to be ungrammatical was ruled out.

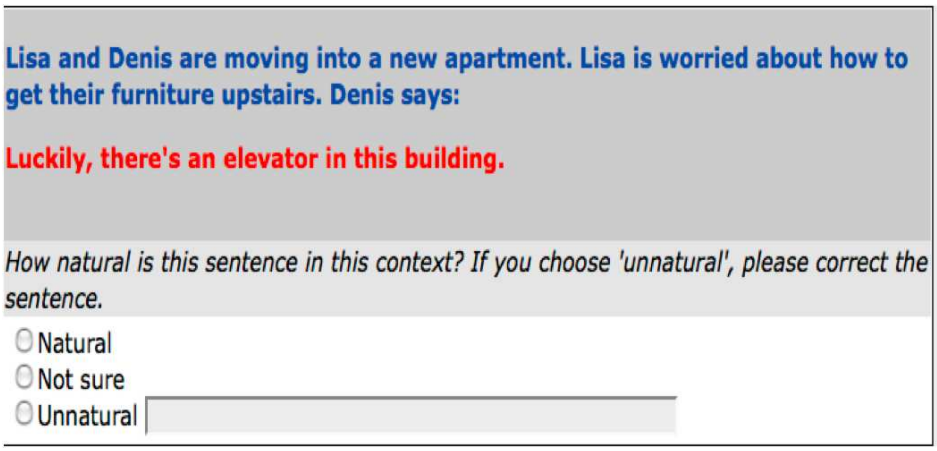
E.g. "There isn't the bowl here."—"here" is included in order to rule out a deictic reading.

Piloting

The last important step before running the big experiment. Try running it on the native speaker controls. They should match your expectations perfectly—if they don't, it suggests a problem with the test.

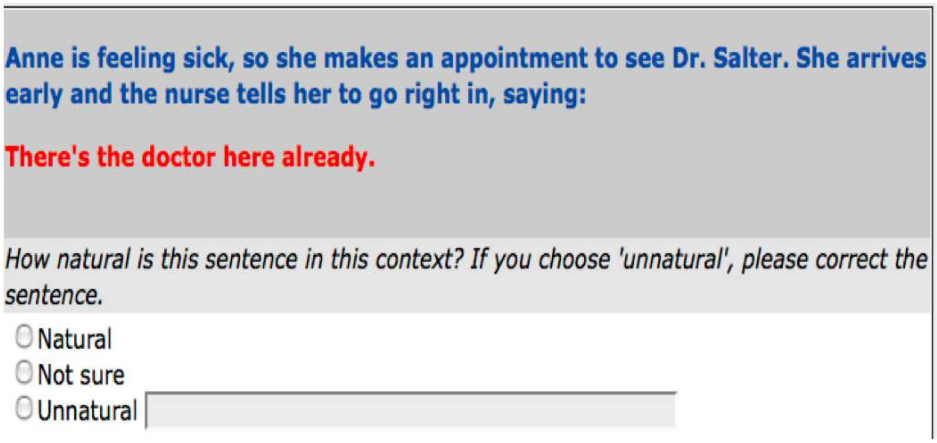
Sample test items

Grammatical *there*-insertion with indefinite article



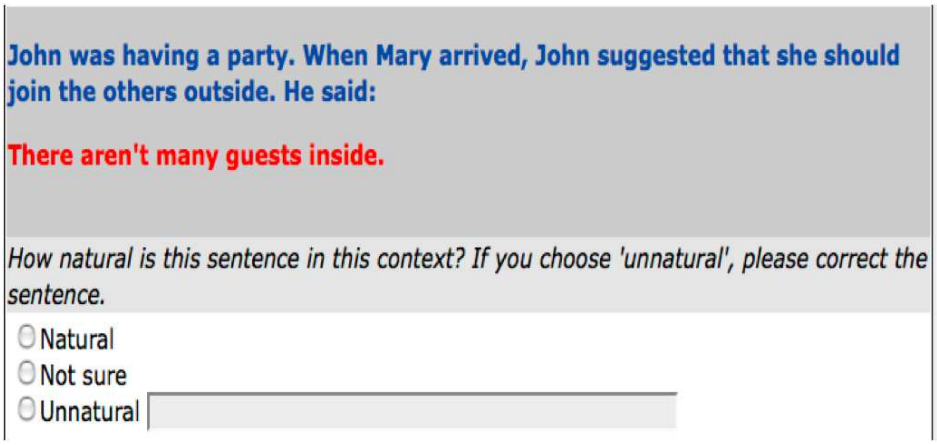
Sample test items

Ungrammatical *there*-insertion with definite article



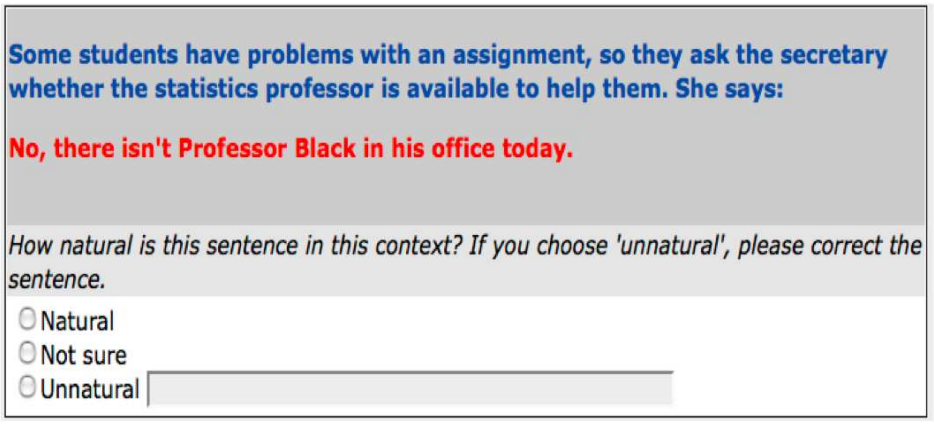
Sample test items

Grammatical *there*-insertion, weak DP, negative



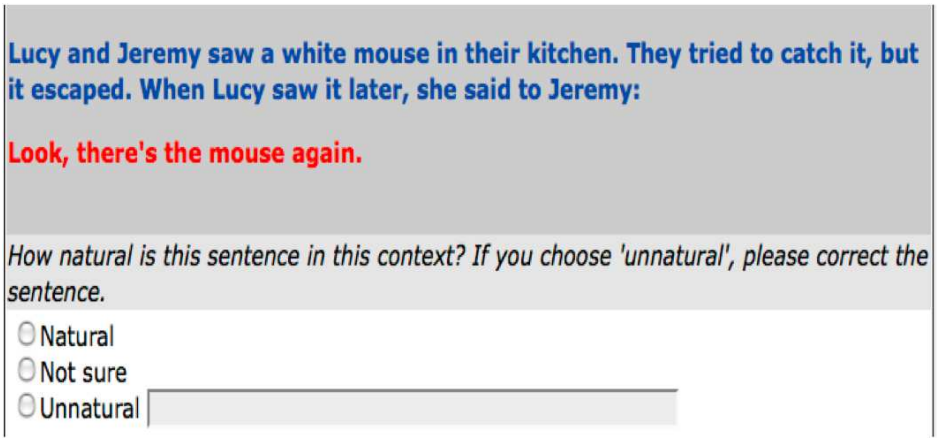
Sample test items

Ungrammatical *there*-insertion, strong DP, negative



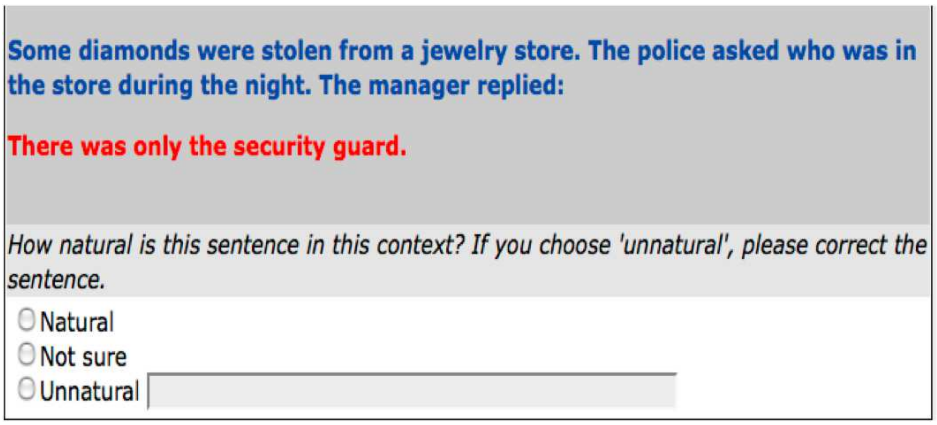
Sample test items

Grammatical deictic with definite article



Sample test items

Grammatical list reading with definite article



Coding: Relevant corrections

There's Mary at home.	⇒	Mary is at home.
There isn't the package here yet.	⇒	The package isn't here yet.
There aren't most guests outside.	⇒	There aren't many guests inside.
There's a suitable bowl over here.	⇒	That bowl over there should do the trick.

Coding: Irrelevant/reverse

There was no one in class today.	⇒	There is no one in class today.
There's the bus at last.	⇒	At last there's the bus.
Ok. There's Mary at home.	⇒	Ok. There's Mary here.
There's a package for you at the post office.	⇒	Just to let you know, there's a package waiting for you at the post office.

Coding: Discard

There's the only copy machine downstairs.	⇒	There's only copy machine downstairs.
Tony seems to be in the garden.	⇒	Is that Tony I hear singing in the garden?
There's another doctor here instead.	⇒	Dr. Smith is covering for Dr. Salter while she is away.
There's a suitable bowl over here.	⇒	That bowl over there should do the trick.

3.4 Experimental Results

In brief: There is a clear difference between grammatical and ungrammatical sentences, for both the intermediate and advanced L2ers: The L2ers know the DE constraint. Moreover, both groups of L2ers reject strong/definite NPs in negative existentials, even though these would be grammatical in the L1. They have not transferred the "Turkish DE" but are showing the DE as appropriate to English.[1]

[1]Statistical information: Our test items included more ungrammatical existentials (definite articles, strong quantifiers and two other types of strong DPs) than grammatical (indefinite articles and weak quantifiers). To enable statistical comparisons, the analysis in Figure 1 is confined to articles and quantifiers.

We conducted 2 factor repeated measures ANOVAs to test for differences, then one factor ANOVAs and post-hoc Scheffé tests, as appropriate, to determine the source of any differences.

Figure 1 (Overall comparison of existential sentences). The 2-factor ANOVA shows a significant effect for group ($f(2, 29)=6.044$, $p<0.01$), for sentence type ($f(3, 29)=370.084$, $p<0.0001$), and a significant interaction ($f(6, 29)=17.896$, $p<0.0001$). Post-hoc Scheffé tests show that the differences between grammatical and ungrammatical are significant for all groups, both for affirmatives and negatives. Comparing grammatical affirmatives and negatives, there are no significant differences for any group. For the ungrammatical sentences, the advanced group and native speakers did not differ significantly, but the intermediate group is significantly more likely to accept ungrammatical sentences than the advanced group and the native speakers (ungrammatical affirmatives: $f(2, 29)=25.504$, $p<0.0001$, ungrammatical negatives: $f(2, 29)=14.218$, $p<0.0001$).

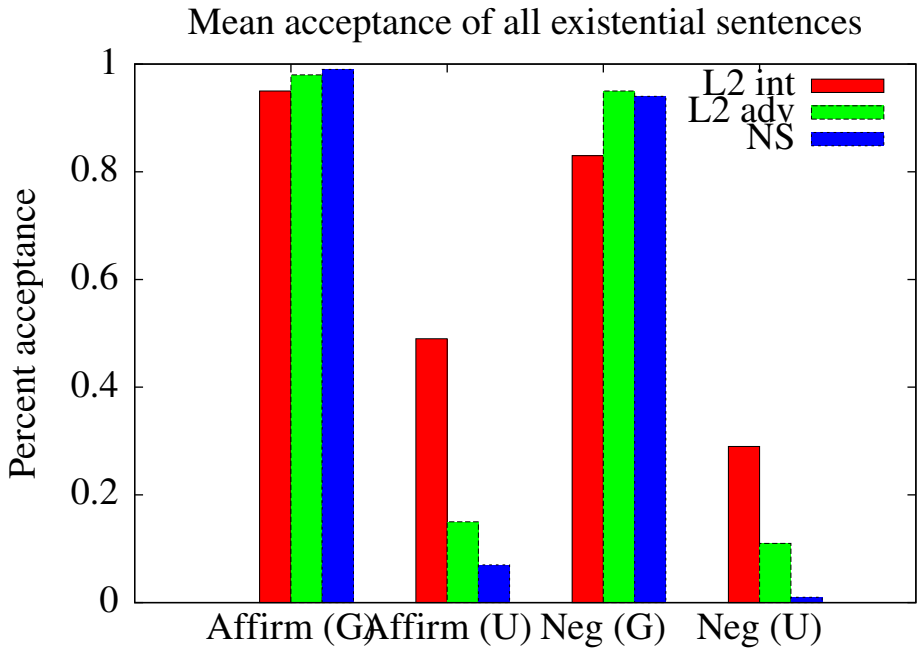
Figure 2 (Ungrammatical affirmative existentials). A 2 factor repeated measures ANOVA shows a significant effect for group ($f(2, 29)=17.048$, $p<0.0001$), a significant effect for sentence type ($f(3, 29)=5.896$, $p<0.001$) and a significant interaction ($f(6, 29)=3.407$, $p<0.005$). The performance of the advanced subjects does not differ from the native speaker controls, whereas the intermediate subjects are significantly more likely than the controls and the advanced group to accept all ungrammatical existentials except for proper names. The intermediate group shows a significant difference across conditions ($f(3, 11)=7.139$, $p<0.001$), with proper names being accepted significantly less than definite articles and strong quantifiers. For the advanced group and the native speakers, there are no significant differences across conditions.

Figure 3 (Ungrammatical negative existentials). A 2 factor repeated measures ANOVA shows a significant effect for group ($f(2, 29)=16.422$, $p<0.0001$), a significant effect for sentence type ($f(3, 29)=7.807$, $p<0.001$) and a borderline significant interaction ($f(6, 29)=2.204$, $p=0.05$). Native speakers reject all four sentence types outright, regardless of subtype, and the advanced L2ers show a similar pattern, with, again, no significant differences between them and the control group. The intermediate group differs from the native speakers on all four sentence types and they differ from the advanced group on possessives and quantifiers. Proper names are more consistently rejected than other ungrammatical cases. The intermediate group shows a significant difference across conditions ($f(3, 11)=4.391$, $p<0.01$), with proper names being accepted significantly less than definite articles. The advanced group also shows a significant difference across conditions ($f(3, 9)=50268$, $p<0.01$), with names and possessives both differing from definite articles. The native speakers show no significant difference across conditions.

Figure 4 (Control items). In list (e.g. (7a)) and deictic (e.g. (7b)) contexts, it is possible for a definite expression to co-occur with there. A 2 factor repeated measures ANOVA (comparing list and deictic readings) shows a significant effect for group ($f(2, 29)=6.695$, $p<0.005$), no effect for sentence type ($f(1, 29)=2.116$, $p=0.1565$) and a borderline significant interaction ($f(2, 29)=3.253$, $p=0.0531$). All subjects accept list readings as being natural. In other words, they are prepared to accept sentences involving there followed by a definite DP, which suggests that they are not simply following a strategy of rejecting all sentences involving definites, something which might otherwise account for their successful performance on the existentials. For the native speakers and the advanced group there is no significant difference between acceptances of list and deictic readings; however, the intermediate subjects are significantly less accepting of deictic readings ($f(1, 11)=5.037$, $p<0.05$).

As for the ungrammatical sentences involving indefinite articles (such as (7c)), these were strongly rejected by the advanced group and the native speakers, and also rejected (to a somewhat lesser extent) by the intermediate group. These results suggest that the L2ers were not simply following a strategy of accepting all sentences involving indefinites, which, again, might otherwise account for their accuracy with existentials. A one factor ANOVA shows that

Fig 1. Overall comparison of existential sentences



there is a significant difference between the groups, with the intermediate group differing from both the other groups ($f(2, 29)=7.236$, $p<0.005$).

Fig 2. Ungrammatical affirmative existentials

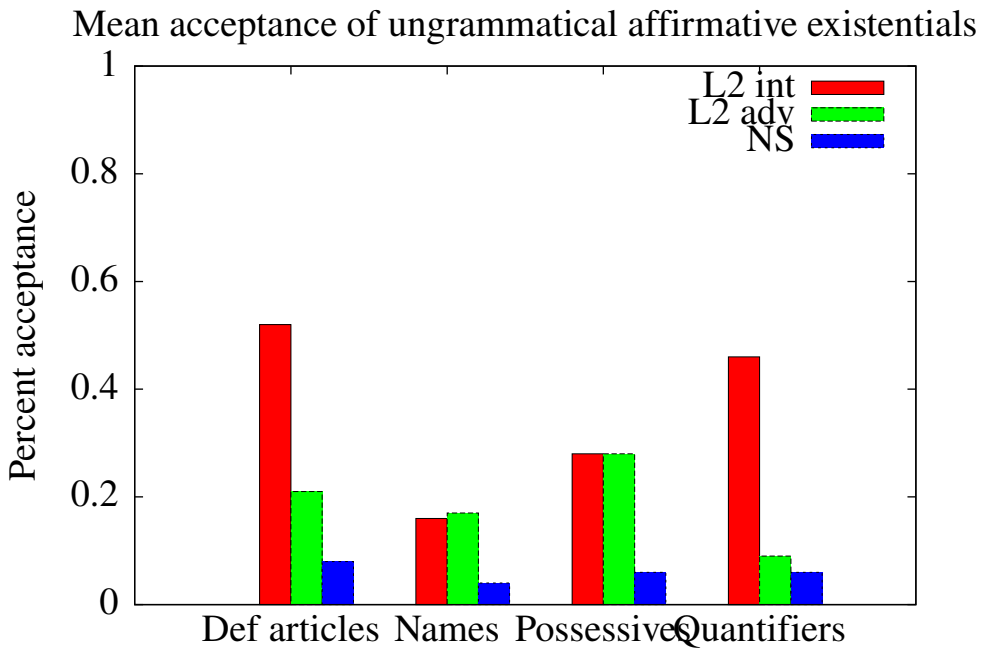


Fig 3. Ungrammatical negative existentials

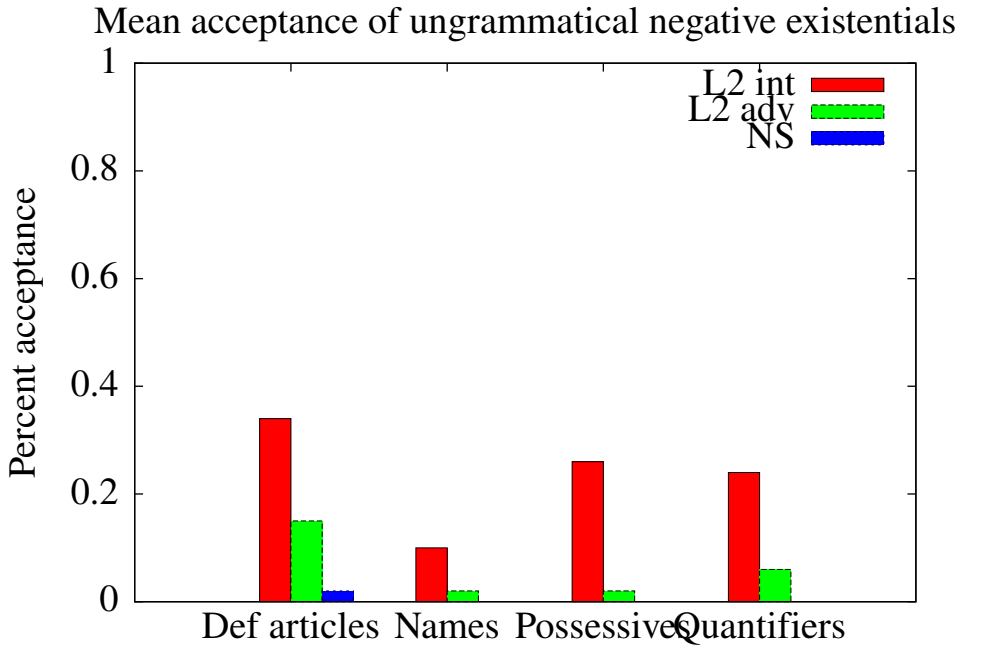
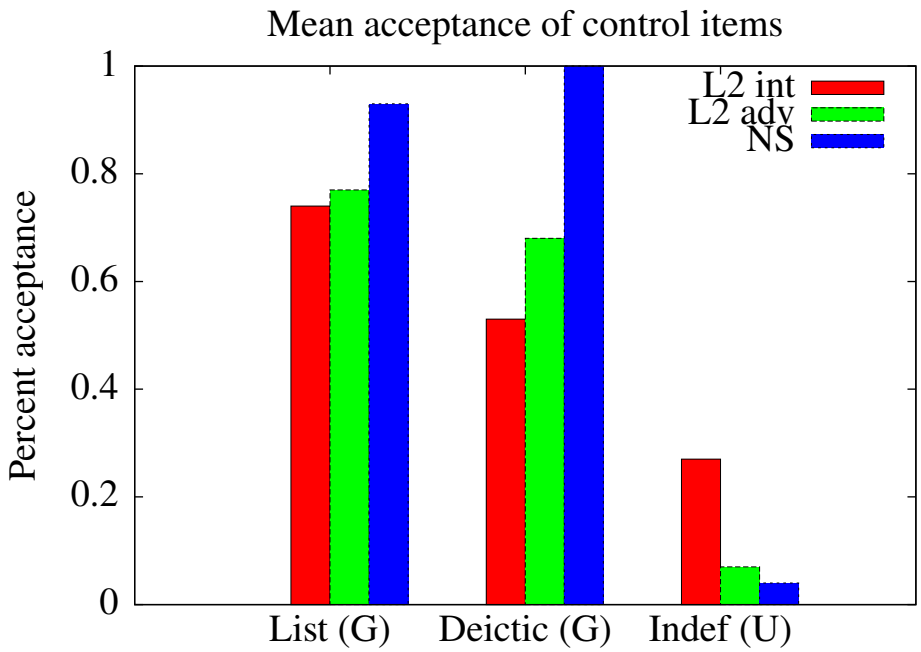


Fig 4. Control items



Summary of results

- L2ers at all levels differentiate between grammatical and ungrammatical cases of *there*-insertion.
- Performance by the advanced group is not noticeably different from native speakers.
- No evidence of transfer: L2ers reject sentence types which are ungrammatical in the L2 and grammatical in the L1.

4 Discussion

4.1 Source of L2 knowledge?

Discussion

The L2ers' knowledge of English seems to include the English version of the DE. Where did they get it?

Frequency? No, L2ers show evidence of knowing how the DE applies, regardless of frequency of various determiners in this construction.

Instruction? (*There is a.../There are some...*). List readings (*There is the...*) were accepted.

Transfer? Not in an obvious way, negative existentials with definite DPs were rejected.

UG? Maybe. But what principles/parameters explain the DE?

References

Enç, Mürvet. 1991. The semantics of specificity. *Linguistic Inquiry* 22(1): 1–25.

Ionin, Tania. 2006. This is definitely specific: Specificity and definiteness in article systems. *Natural Language Semantics* 14: 175–234.

Kornfilt, Jaklin. 1997. *Turkish*. Routledge.

Lardiere, Donna. 2005. On morphological competence. In Laurent Dekydtspotter, Rex A. Sprouse & Audrey Liljestrand (eds.) *Proceedings of the 7th Generative Approaches to Second Language Acquisition Conference (GASLA 2004)*, 178–192. Somerville, MA: Cascadilla Proceedings Project.

Milsark, Gary. 1977. Toward an explanation of certain peculiarities of the existential construction in English. *Linguistic Analysis* 3(1): 1–29.

White, Lydia. 2003. Fossilization in steady state L2 grammars: persistent problems with inflectional morphology grammars: persistent problems with inflectional morphology. *Bilingualism: Language and Cognition* 6(2): 129–141.

White, Lydia. 2008. Definiteness effects in the L2 English of Mandarin and Turkish speakers. In Harvey Chan, Heather Jacob & Enkeleida Kapia (eds.) *BUCLD 32: Proceedings of the 32nd annual Boston University Conference on Language Development*, 550–561. Somerville, MA: Cascadilla Press.

White, Lydia, Alyona Belikova, Paul Hagstrom, Tanja Kupisch & Öner Özçelik. 2009. Restrictions on definiteness in L2 English. In Jane Chandlee, Michelle Franchini, Sandy Lord & Gudron-Marion Reiner (eds.) *BUCLD 33: Proceedings of the 33rd annual Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press.